(12)

NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY

SUBMARINE BASE, GROTON, CONN.

AD-A163 894

REPORT NUMBER 1068

AN EVALUATION OF COLOR SETS FOR CRT DISPLAYS

by

David F. Neri
Alan R. Jacobsen
and
S. M. Luria

DTIC
ELECTE
FEB 1 1 1986
S D
D

Naval Medical Research and Development Command
Research Work Unit M0100.001-1022

FILE COPY

REPRODUCED FROM
BEST AVAILABLE COPY

Released by:

C. A. Harvey, CAPT, MC, USN
Commanding Officer
Naval Submarine Medical Research Laboratory
31 December 1985

6 2 1 0 105

Approved for public release; distribution unlimited

AN EVALUATION OF COLOR SETS FOR CRT DISPLAYS

by

David F. Neri, LT, MSC, USNR
Alan R. Jacobsen, Ph.D.
S. M. Luria, Ph.D.

NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY
REPORT NUMBER 1068

NAVAL MEDICAL RESEARCH AND DEVELOPMENT COMMAND
Research Work Unit M0100.001-1022

Approved and Released by

C. A. Harvey

C. A. Harvey, CAPT, MC, USN
Commanding Officer
NAVSUBMEDRSCHLAB

Approved for public release; distribution unlimited

SUMMARY PAGE

PROBLEM

To determine whether, and why, some color sets patterned after those recommended by investigators for use in visual displays are better than others in facilitating a color matching task similar to that which operators would perform using color-coded CRT displays on submarines.

FINDINGS

The color matching task was performed much more quickly and accurately with some color sets than with others. The larger the amount of estimated color difference between the two most similar colors in a set, the better the performance.

APPLICATION

The best of a representative sample of color sets for use with a color matching task have been identified, and it has been shown that, in general, the smallest perceived color difference between any two colors in a set should be maximized.

ADMINISTRATIVE INFORMATION

This research was conducted as part of the Naval Medical Research and Development Command Work Unit M0100.001-1022 - "Enhanced performance with visual sonar displays." It was submitted for review on 25 Oct 1985, approved for publication on 31 Dec 85, and designated as NSMRL Report No. 1068.

PUBLISHED BY THE NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY

Abstract

Ten subjects performed a color matching task with ten sets of seven colors, adapted from sets recommended in the literature. The color sets covered a wide range in color difference values (ΔE^* in CIELUV, 1976). Performance with some color sets was significantly better than that with others on a task where color discrimination was important and the colors needed to be discriminated quickly and accurately. Effectiveness of a color set did not depend on the inclusion of any particular colors; rather, the relationships among colors proved to be important. The larger the color difference value between the two most perceptually similar members of a set, the shorter the reaction time and the fewer the errors on the color matching task.

Color-coding is being added to visual displays on submarines to improve the ability of operators to interpret information. When this is done, the first problem which must be solved is which colors to use. While most researchers appear to agree that the choice of proper colors is task-specific, several have recommended sets as containing clearly discriminable colors which they believe will result in the least amount of confusion, and the most improvement in performance. There is rough agreement on which colors should be used, but a certain amount of variability in these recommendations remains. It is not certain how much these differences will affect performance.

A second problem is that few of these recommendations have been based on studies using the cathode ray tubes (CRTs) on which the displays will be produced and which the operators will monitor. It is not at all certain to what extent results obtained with reflected or projected color stimuli can be generalized to CRT displays (Jones, 1962; Laycock, 1982). Studies using actual CRTs are necessary.

Finally, the evaluation of sets of colors has often involved arduous and time-consuming procedures. Butler & McKemie (1974), for example, carried out a series of paired comparisons to ascertain which of a set of 46 colors were the brightest and most conspicuous; this required an enormous amount of work. A highly efficient procedure which should serve the purpose was used by Luria, Neri, & Jacobsen (in press). It involved the measurement of the speed and accuracy with which observers could match a test color from a displayed set of colors to the correct color in that set. Although how good a set of colors is may depend on the specific task which must be performed using those colors, this method provides a good indication of how difficult it is to discriminate among the colors in the set and which colors are posing the most difficulty. Sets posing few such difficulties are, in all probability, suitable for color-coding most display formats.

In this investigation, we have used the matching task described above to evaluate ten sets of seven colors. Many of these sets are patterned after ones which have been recommended in the literature, including two of ours. The major modification made to these sets stemmed from consideration of the color difference value, ΔE^*. This value, part of the 1976 CIELUV system, is an estimate of the perceptual color difference between any two colors of known chromaticity and luminance, with larger values corresponding to larger differences. Although it is not free of some possible theoretical problems, the color space and equations on which this color difference formulation is based "currently offer the most empirically sound foundations for predicting effective color display performance" (Silverstein &

Merrifield, 1985). This estimate of color difference has also been shown to be a good predictor of visual search performance (Carter & Carter, 1981). Therefore, it is of interest to examine whether ΔE^* plays a role in explaining any differences in performance on a color matching and discrimination task.

There are two alternate ways of manipulating color differences between set members: a) varying the mean (or sum) of the differences between colors, and b) varying the minimum difference between any two colors (Carter & Carter, 1982). In this study, the color sets were altered so as to produce a wide range in both the mean and minimum ΔE^* values for the sets. These two independent variables could then be examined for their relative effectiveness in predicting performance. As Carter & Carter (1982) pointed out, maximizing the minimum color difference would seem more appropriate since "a change of a large distance between colors has little or no effect on human performance, but the same change of a small distance has a greater effect" (p. 2937). Consequently, we predicted that if color difference was shown to be related to color set performance, it would be the minimum, rather than mean, color difference that would predict best.

Method

Observers

Ten Navy personnel awaiting enrollment in Submarine School served as voluntary observers. All had normal color vision as determined by the American Optical Hardy-Rand-Rittler Pseudoisochromatic Plates. Those who normally wore corrective lenses did so during the experiment.

Apparatus

The colors were presented on an Advanced Electronics Design Model 512 Color Graphics and Imaging Terminal, driven by a PDP 11/34 Laboratory Computer. The stimuli for all tasks consisted of ten circles in the arrangement of a telephone keypad: a 3 x 3 matrix with the tenth circle centered below. The dimensions of this stimulus arrangement are given in Fig. 1. The first seven circles were used as stimuli and could either contain letters or be filled with white or other colors, depending on the task. The three remaining circles were unfilled for all tasks. Table IV in the Appendix gives a complete description of all colored and white stimuli. The CRT background was 0.14 C/m^2.

Responses were recorded by means of a telephone-style keypad with push-buttons in the same physical arrangement as the circular stimuli. Observers were seated approximately 50 cm from

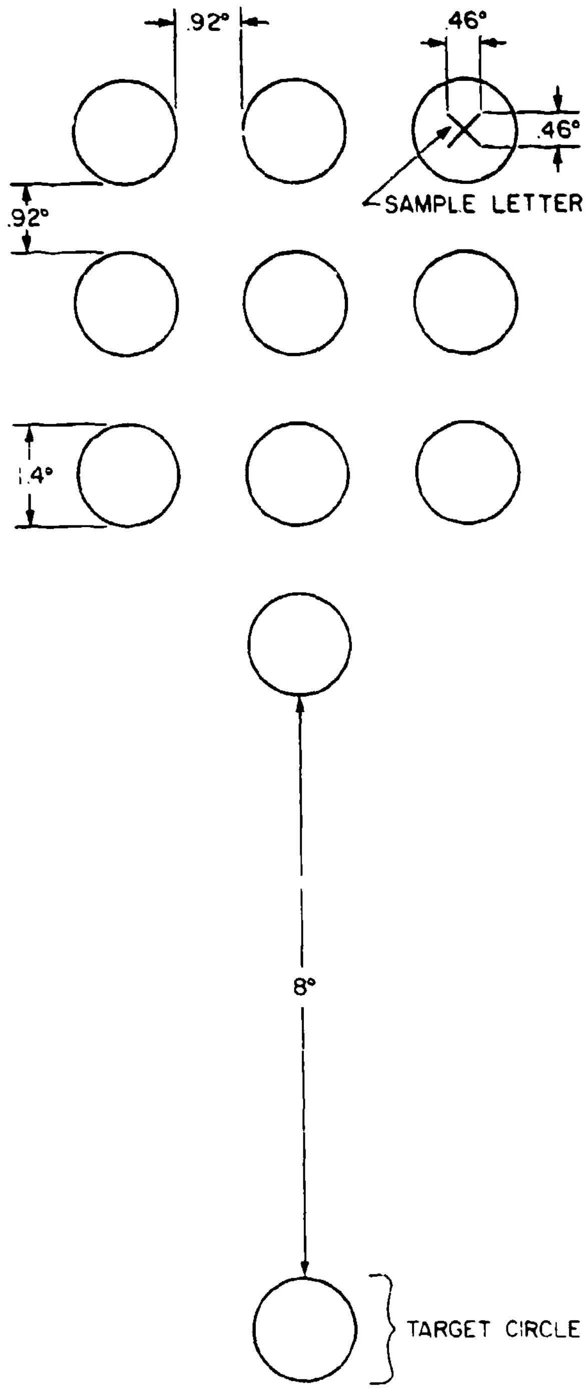

Fig. 1. The stimulus arrangement used in this study. Measurements are in degrees of visual angle at a viewing distance of 50 cm. The figure is drawn to scale.

the CRT screen placed at eye level, with their preferred hand resting on the keypad. Two cool-white fluorescent light bulbs, about 1 m above and behind the observer and covered with neutral density filters, cast 2.7 lux of illumination on the screen, the maximal amount typically found in submarine sonar shacks under operational conditions (Kinney, Luria, Neri, Kindness, & Schlichting, 1981).

Color Sets

The sources that inspired our sets are listed in Table I. Using the brightest white that the CRT could produce as the luminance standard, and D65 as the standard illuminant (Carter & Carter, 1983), color difference values were computed for every possible color pair in each of the ten color sets. The minimum and mean ΔE^* values for the sets are listed in the table. Details of the ways in which we altered the sets to achieve a range of minimum and mean ΔE^* values, the chromaticities and luminances of the colors, and the method by which we reproduced these colors are all given in the Appendix.

TABLE I

NAMES, MINIMUM & MEAN ΔE^* VALUES, AND SOURCES OF COLOR SETS

SET	MINIMUM ΔE^*	MEAN ΔE^*	SOURCE
A	49.1	114.3	Luria et al., 1985
B	47.4	137.1	Laycock, 1982
C	42.4	89.5	Jacobsen, 1985a
D	35.6	115.1	Meister & Sullivan, 1969
E	34.3	121.4	Laycock, 1982
F	33.9	140.8	Meister & Sullivan, 1969
G	24.3	67.3	Feallock et al., 1966
H	9.3	33.1	Halsey & Chapanis, 1951
I	4.5	35.1	Butler & McKemie, 1974
J	2.5	39.5	Butler & McKemie, 1974

Procedure

Observers first completed the following exercise to familiarize themselves with the experimental procedure and reduce the possibility of a practice effect. The computer presented a stimulus set consisting of the ten circles with different letters of the alphabet displayed inside the first seven, starting in the upper left and proceeding to the right and then down. This stimulus set was visible for an entire block of trials. The observer studied the locations of the letters as long as he desired, in order to learn their arrangement. The computer then randomly presented one of these as a target letter, 1 s after a warning tone and 8 degrees of visual angle below the middle column of circles. The observer's task was to press the button on the keypad corresponding to the location of the target letter in the stimulus set, as quickly as possible but without sacrificing accuracy. The keypress caused the letter to disappear, and the computer recorded response time (RT) and accuracy. After 2 s there was another warning tone and another letter was presented in the same position. This procedure continued, using randomization without replacement, until all seven letters were responded to correctly once. After seven presentations resulting in correct responses, the computer displayed the average RT for these responses.

Observers repeated this procedure several times. On each block of trials the locations of the seven letters in the keypad remained the same but the order of presentation of the target letters was re-randomized. Observers completed enough blocks to cause their average RTs to reach level off -- usually five to six blocks.

After the practice session, the observers completed the main experimental task with each of the ten color sets. The procedure was exactly analogous to that of the practice exercise, with the following exceptions. The seven circles were color-filled, and one of the colored circles appeared as the target circle below the keypad arrangement 0.75 s after a warning tone with a 1.3 s delay between trials.

For each color set a completed block of trials required correct responses to ten presentations of each of the seven colors. There was thus a total of 70 correct responses in addition to any incorrect responses. The positions of the colors within the display were randomized across observers. The order of presentation of the target colors was randomized separately for each observer and each color set. Again the computer recorded RT for each correct response. Incorrect responses were not counted toward the 70 trial criterion, but information on the color presented, its position in the display, the button pressed by the observer, and the color in the display corresponding to this button position were all recorded.

It took approximately 3 min. to complete 70 correct trials for a color set. All ten sets were given in a single session. The order of presentation of the ten sets was counterbalanced across the ten observers so that each set occurred in each presentation position once. After the first five sets, observers were given a ten minute rest period followed by a control set and the remaining five color sets.

The procedure used with the control set was similar to that used with the experimental one. Again, the ten circles were displayed. The first seven were outlined in white, denoting those to be attended to, the last three in red. One of the seven white circles was filled with white 0.75 s after a warning tone. The observers again responded by pressing the keypad button corresponding to the location of the filled circle. There was a 1.3 s delay between trials. Ten correct responses per position, for a total of 70, were again required for completion of the task. This control set thus provided RTs to the same visual display using the same response panel as in the experimental task, but without including the time to compare the colors and decide on a match. An entire session, including practice, took approximately 1 1/2 hrs.

Results

Differences in Reaction Time Among Color Sets

The average RTs to the white circles of the control set were subtracted from the average RTs to the ten color sets in each of the seven positions for each subject. Fig. 2 shows that RT to the white circles varied with position ($F(6,54)=7.12$, $p<.01$), as found previously (Luria, et al., in press). The resulting data using the color sets, therefore, represent only the times to perceptually match the colored stimuli and decide on a response, without the motor component.

The mean RTs to the various color sets were significantly different, according to a one-way repeated measures ANOVA, ($F(9,81)=8.37$, $p<.01$). The Newman-Keuls post-hoc test showed that set H was responded to slower than all other sets except J. Set J, in turn, was responded to slower than all other sets except H. Set I was slower than C. No other differences were significant. The mean RTs are shown as the solid bars in Fig. 3.

Differences in Errors Among Color Sets

The differences in the number of matching errors made among the ten color sets were also statistically significant, according to a similar analysis ($F(9,81)=8.79$, $p<.01$). The mean errors for the color sets are shown as the open bars in Fig. 3. Newman-Keuls tests revealed that set H yielded more errors than all but I and J. Set J resulted in more errors than all sets except H. Thus, both the RT and error analyses showed that sets H, I, and J contained colors more difficult to distinguish than those of most of the other sets. In fact, the strong agreement between RTs and errors is evidenced by the high linear

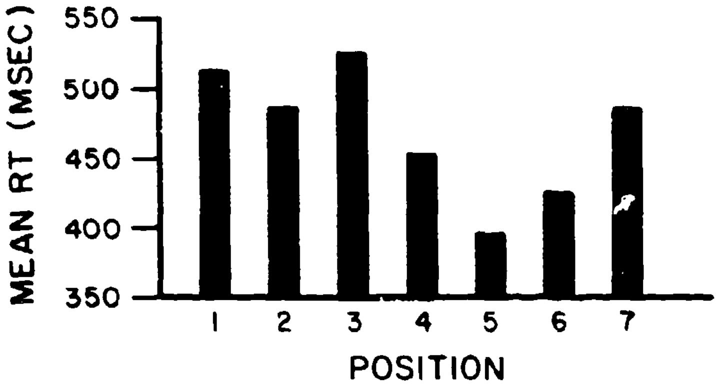

Fig. 2. The average reaction times to the white circles used in the control study. Positions 1 through 7 refer to the first seven positions in the display in Fig. 1.

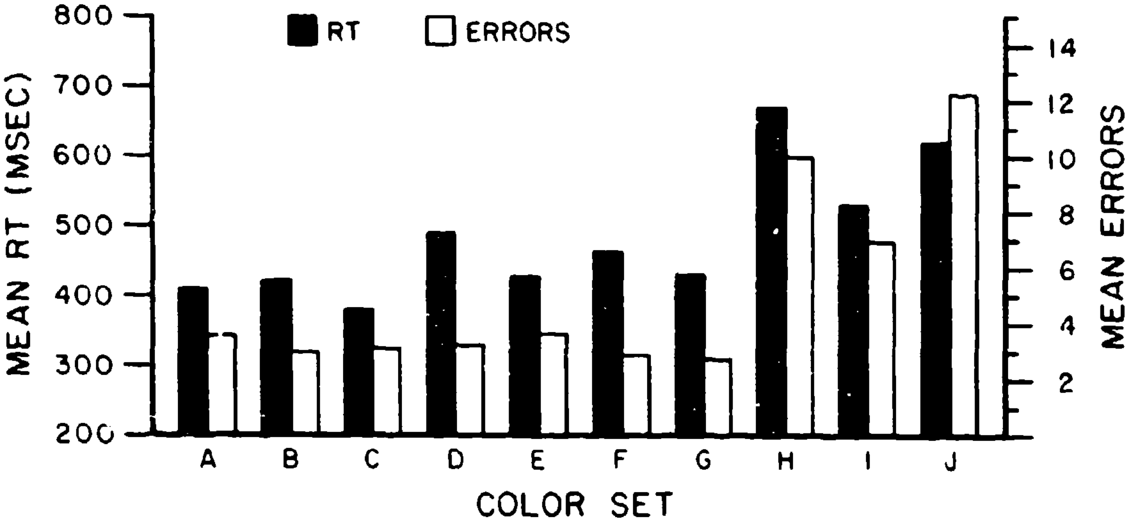

Fig. 3. The average reaction times (solid bars and left axis), and the average number of errors (open bars and right axis) for the ten color sets compared in this study. Color sets are arranged in order of decreasing ΔE^* values.

correlation between the two (r=0.91, p<.01).

Differences in Reaction Time Among Colors in Each Set

We examined the pattern of differences in the corrected RTs among the colors comprising the sets. Separate ANOVAs for each set revealed that seven of the ten sets showed significant differences in RTs between colors. Sets B, H, I, and J were significantly different at the .01 level, with A, D, and E significantly different at the .05 level. Sets C, F, and G showed no differences between colors. Table V in the Appendix shows the results of the Newman-Keuls tests for the seven sets with significant differences among colors. Table II lists the colors from fastest to slowest for each of the ten sets. This table shows that there is no "good" or "bad" color. How fast a color is responded to depends on the other colors in the set. The same or similar colors can even be the fastest in one set and the slowest in another, as gray is in sets D and E.

TABLE II

THE COLORS OF THE SEVEN SETS WITH SIGNIFICANT DIFFERENCES AMONG THEM, LISTED FROM FASTEST TO SLOWEST

	COLOR SET						
	A	B	D	E	H	I	J
F A	Red	Blue	Blue	Gray	Gre-Yel	Yel-Gre	Yel-Whi
S T	Dk Blue	Yel-Gre	Red	Yel-Gre	Yellow	Yellow	Blue
. .	White	White	Yellow	Orange	Blue	Cyan	Yel-Gre
. .	Aqua	Orange	Magenta	Purple	Red-Ora	Magenta	Orange
. S	Yellow	Red	Green	Red	Mauve	Pur-Blu	Cyan 3
L O	Pink	Purple	Amber	Blue	Purple	Y-G-Whi	Cyan 1
W	Purple	Cyan	Gray	Cyan	Red	White	Cyan 2

A preliminary way to look at these differences among colors is to examine the range of luminances found in each color set. These ranges themselves have a range of 2 log units; the colors differ by as much as 222.7 C/m^2 in set A, but only 2.1 C/m^2 in set H. Luminance differences would be expected to affect performance, and they may well have, since luminance range was inversely correlated with mean RT (r=-0.71, p<.05). Thus the greater the luminance range in a set, the shorter the average RT. However, the correlation with mean errors, while in the same direction, fell short of significance (r=-0.59, n.s.).

Relationship of Color Differences to RT and Errors

A more complete estimate of color difference is ΔE^*, which incorporates chromaticity differences as well as the luminance differences referred to above. Two separate regression analyses were computed to examine the effects of ΔE^* on RT and errors. With RT as the dependent measure, and mean and minimum ΔE^* as the independent measures, R=.841 (p<.05). Thus some 71% (R^2) of the variance in RT is linearly accounted for by mean and minimum E^* considered together. However, mean and minimum ΔE^* correlate very highly with each other (r=.875, p<.01). The question arises, therefore, do they contribute equally to the explanation of RT differences, or is one more important than the other?

Further analyses revealed the relative contributions of mean and minimum ΔE^* measures. The simple correlation between minimum ΔE^* and RT is -.840 (p<.01). The trivial difference between the absolute value of -.840 and an R of .841 shows that incorporation of mean ΔE^* in the regression adds virtually nothing to minimum ΔE^*'s already impressive estimation of RT. This is made even more explicit from examination of the semipartial correlation coefficients. With minimum ΔE^* in the regression equation, the increase in R^2 which occurs when mean ΔE^* is added is only .001. That is, once minimum ΔE^* is taken into account, mean ΔE^* accounts for only an additional 0.1% of the variance in RT. Conversely, with mean ΔE^* accounted for, adding minimum ΔE^* accounts for an additional 19.0% of the RT variance. We can conclude that mean ΔE^* has no unique relationship to RT, i.e. no relationship beyond what can be accounted for by minimum ΔE^*. On the other hand, minimum ΔE^* is uniquely related to RT, and to RT holding mean ΔE^* constant.

With number of errors as the dependent variable, R=.846. Therefore, once again a very large portion of error variance (72%) is linearly accounted for by both mean and minimum ΔE^*. However, as above with RT, mean and minimum ΔE^* do not contribute equally to explaining error variance. The semipartial correlation coefficients reveal that once minimum ΔE^* is in the equation, the addition of mean ΔE^* accounts only for an additional 2% of error variance. However, with mean ΔE^* in the regression, minimum ΔE^* explains an additional 9%. The high simple correlations between mean ΔE^* and RT (r=-.72, p<.05), and mean ΔE^* and errors (r=-.79, p<.01) exist because information carried by mean ΔE^* is highly redundant with that carried by minimum ΔE^*, and it is minimum ΔE^* that is doing most of the work in explaining variations in RT and errors.

Closer examination of these relationships between performance measures and ΔE^* measures reveals that there may be an upper limit to the increase in performance that is realizeable as a result of maximizing the minimum ΔE^* value. Figs. 4 and 5 show RT and errors, respectively, as a function of mean and minimum ΔE^* values for the ten color sets. It can be seen that, at low values of mean and minimum ΔE^* (<45), minimum ΔE^* is more linearly related to performance than mean ΔE^*. At larger values (>45), the two measures show signs of coinciding. It appears that performance may increase up to about 40 or 50 units and then level off. The more important measure of minimum ΔE^* shows signs of flattening at about 40-45 ΔE^* units while the mean ΔE^* values, although highly variable below 40, also show signs of stability above 40.

Color Confusions

Confusion matrices were constructed for each set to uncover confusion patterns among colors. There were 12 pairs of colors that resulted in more than ten errors, as shown in Table III. The three color sets that are not represented had no color pairs with ten or more errors and were also those with the least number of overall errors. Note that the mean ratio of errors was about 5:1 in favor of a particular direction of confusion. In the first case shown, pink was erroneously matched to white 18 times but white was erroneously matched to pink only 4 times. In another case 14 errors were in one direction and none in the opposite. This finding was not due to only one or two subjects making large numbers of unidirectional errors. By and large, the same pattern is found when the subjects are considered individually. A possible explanation is given in the Discussion.

Discussion

Our color matching procedure effectively divided the ten sets into two groups. The better sets (A, B, C, D, E, F and G), although not significantly different from each other, are certainly superior to those in the second group (sets H, I and J). Thus, where color discrimination is important and colors need to be discriminated quickly and accurately, not all color sets are equally good. We believe that for these types of applications, sets A through G would be appropriate choices. This is an important result, because the matching procedure directly measures how difficult it is to distinguish among the colors in a set, surely a critical consideration in most applications of color to visual displays.

Given the differences among sets, what makes one set of colors better than another? It does not appear to depend on the inclusion of a few particular colors. As can be seen in Table

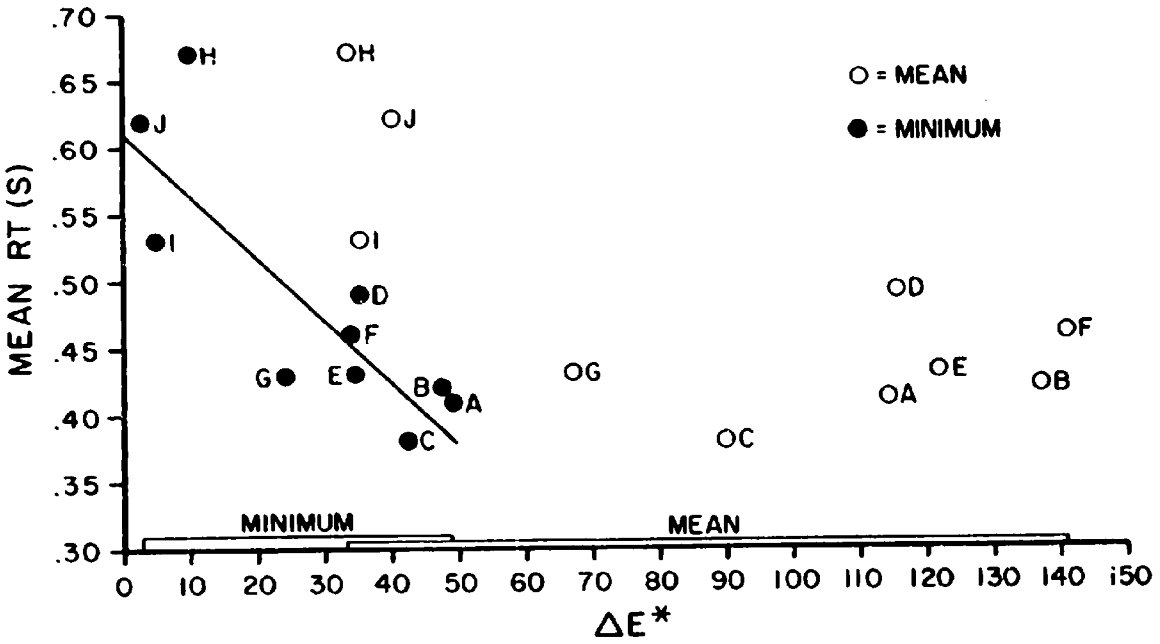

Fig. 4. Average reaction times for the ten color sets plotted both in terms of their minimum and mean ΔE^* values. The brackets along the horizontal axis indicate the ranges of minimum and mean values. The straight line is a least-squares fit to the minimum ΔE^* data points.

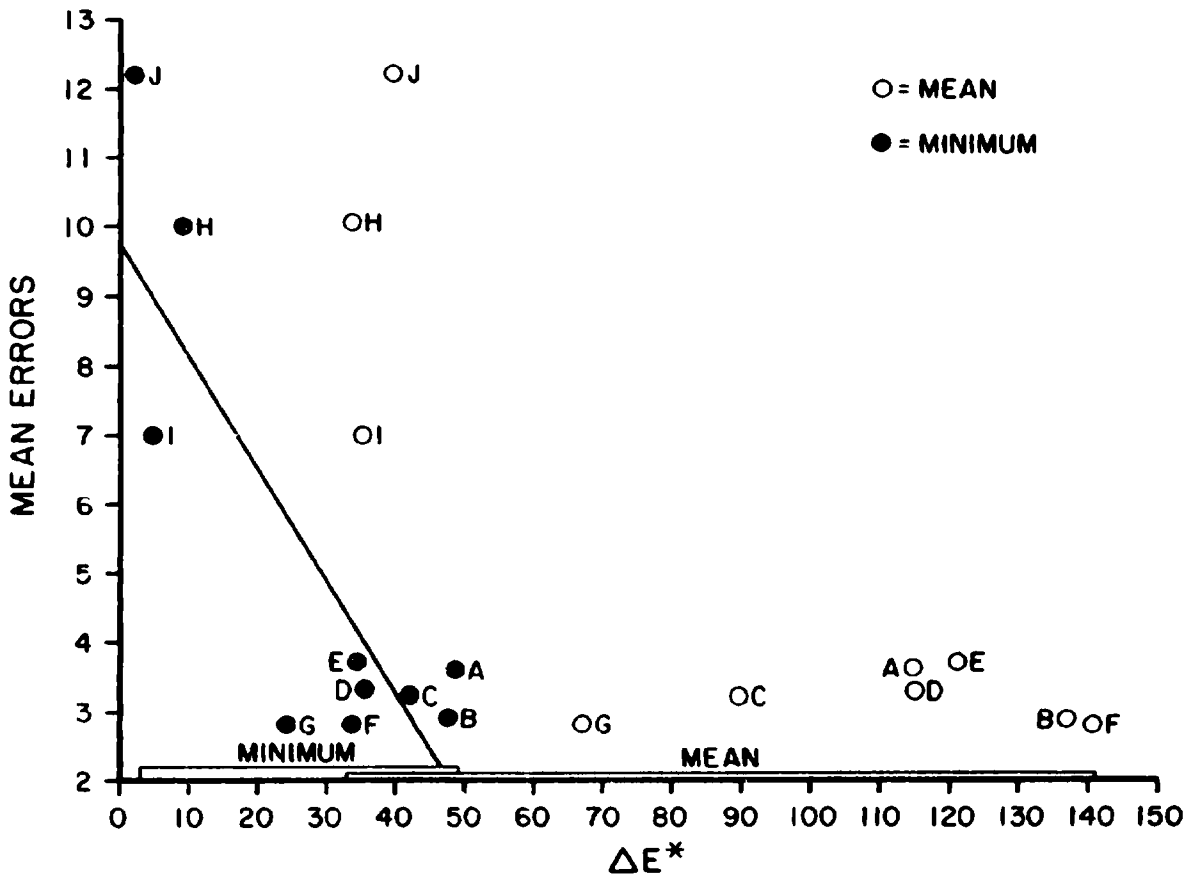

Fig. 5. Average errors for the ten color sets plotted both in terms of their minimum and mean ΔE^* values. The remainder of the figure is the same as Fig. 4.

TABLE III

COLOR PAIRS RESULTING IN A TOTAL OF MORE THAN TEN ERRORS.

COLOR SET	COLOR PRESENTED	RESPONSE	TRIALS
A	Pink	White	18
	White	Pink	4
C	Tan	White	14
	White	Tan	0
D	Gray	Magenta	8
	Magenta	Gray	4
E	Cyan	Blue	10
	Blue	Cyan	6
H	Blue	Purple	24
	Purple	Blue	16
	Red-Ora	Yellow	10
	Yellow	Red-Ora	3
	Red	Mauve	27
	Mauve	Red	7
I	Magenta	Pur-Blue	20
	Pur-Blue	Magenta	17
	White	Cyan	11
	Cyan	White	1
J	Orange	Yel-Gre	12
	Yel-Gre	Orange	4
	Cyan 1	Cyan 2	40
	Cyan 2	Cyan 1	4
	Cyan 2	Cyan 3	37
	Cyan 3	Cyan 2	13

II, colors that resulted in fast RTs in one set often resulted in slow RTs in another set. There was no pattern that emerged to suggest that blue, red or green, for example, should always be included. What appear more important are the relationships among the colors, or how different each color is from every other one. Furthermore, the way in which differences among colors in a set are estimated is critical. The mean (or, equivalently, the sum) ΔE^* value for a set of colors is not uniquely related to performance on our color matching task. The minimum ΔE^* value for a color set is uniquely related to performance.

These findings are particularly interesting in light of the results of Carter & Carter (1981) and Jacobsen (in press). The former showed that both mean search time and relative fixation rate in a search task stabilized at about 40 units of ΔE^*. Above this value, neither measure of performance improved, and below it both measures deteriorated. These authors suggested that the number of colors to be used in an application be kept to a value such that the minimum color distance between any two remains above 40 units. Jacobsen (in press) showed that, with a black background, the number of errors on a color recognition task continued to decrease as the minimum ΔE^* value for individual colors increased, until a value of about 40 was reached. Larger color differences did not decrease the number of errors any farther. The present results, as far as they go, are consistent with these findings. Unfortunately, the size of the minimum ΔE^* values did not exceed 50, so a strong statement about a critical value of 40 or 45 ΔE^* units cannot be made based on our results alone. However it is safe to conclude that minimum ΔE^* appears to hold up well as a predictor of performance even for relatively large color differences. The display designer would do well to attempt to maximize the minimal ΔE^* differences among colors in order to optimize performance on tasks which rely on color discrimination. Fortunately, there is now a computer program to do just that (Silverstein, Lepkowski, Carter, & Carter, 1986).

Finally, the asymmetrical color confusions shown in Table III at first seem quite puzzling. They are not accountable by ΔE^* measures, which are bi-directional. The unidirectionality is probably due to the effect of surround brightness on the perception of the color and is probably an artifact of the display design. As Fig. 1 shows, the colored circles in the keypad display are spaced less than 1 deg apart. The target circle is a full 8 degrees below the lowest keypad circle. The circles in the keypad display are grouped with other luminous stimuli around them while the target circle is alone. Since the surrounding raster is quite dim (0.14 C/m^2), the simultaneous brightness contrast phenomenon would predict that the target circles would appear lighter than their counterparts in the keypad display. In fact, for 9 of the 12 confusion pairs in

Table III, it was a brighter color similar in hue that was chosen as the response. The importance of background has already been shown in a different task (Jacobsen, in press). Further experimentation explicitly examining a range of background luminances on the same task as used in this study is currently underway.

Acknowledgement

The authors thank Dr. James A. Worthey for writing some of the computer programs used in the approximation of colors and Dr. William H. Rogers for his insightful comments on an earlier draft of this manuscript.

References

Butler, W. B., & McKemie, W. M. (1974). Engineering guidelines for the use of color in a sonar display (Document No. 773-AU-9550-U). Austin, TX: Tracor, Inc.

Carter, E. C., & Carter, R. C. (1981). Color and conspicuousness. Journal of the Optical Society of America 71(6), 723-729.

Carter, R. C., & Carter, E. C. (1982). High-contrast sets of colors. Applied Optics 21(16), 2936-2939.

Carter, R. C., & Carter, E. C. (1983). CIE $L^*u^*v^*$ color-difference equations for self-luminous displays. Color Research and Application 8(4), 252-253.

Cook, T. C. (1974). Color-coding - A review of the literature (Technical Note 9-74). Aberdeen Proving Ground, MD: Human Engineering Laboratory. (NTIS No. AD A-001 555).

Feallock, J. B., Southard, J. F., Kobayashi, M., & Howell, W. C. (1966). Absolute judgment of colors in the Federal Standards System. Journal of Applied Psychology 50(3), 266-272.

Granville, W. C., & Nickerson, D. (1943). Trichromatic specifications of intermediate and special colors of the Munsell system. Journal of the Optical Society of America 33, 376-385.

Halsey, R. M., & Chapanis, A. (1951). On the number of absolutely identifiable spectral hues. Journal of the Optical Society of America 41, 1057-1058.

Jacobsen, A. R. (1985). The effect of set size on time to recall color coded information (Report No. 1044). Groton, CT: Naval Submarine Medical Research Laboratory.

Jacobsen, A. R. (in press). The effect of background luminance on color recognition. Color Research and Application.

Jones, M. R. (1962). Color coding. Human Factors 4, 355-365.

Kelly, K. L., Gibson, K. S., & Nickerson, D. (1943). Tristimulus specification of the Munsell Book of Color from spectrophotometric measurements. Journal of the Optical Society of America 33, 355-375.

Kinney, J. A. S., Luria, S. M., Neri, D. F., Kindness, S. W., & Schlichting, C. S. (1981). Surveys of lighting and working conditions in submarine sonar shacks (Report No. 955). Groton, CT: Naval Submarine Medical Research Laboratory.

Laycock, J. (1982). Recommended colours for use on airborne displays (Technical Report No. 82110). Farnborough, Hants, UK: Royal Aircraft Establishment.

Luria, S. M., Neri, D. F., & Jacobsen, A. R. (in press). The effects of set size on color matching using CRT displays. Human Factors.

Meister, D., & Sullivan, D. J. (1969). Guide to human engineering design for visual displays. Washington, DC: Office of Naval Research. (NTIS No. AD 693 237).

Neri, D. F., Luria, S. M., & Kobus, D. A. (in press). The detection of various color combinations under different chromatic ambient illuminations. Aviation, Space, and Environmental Medicine.

Nickerson, D., Tomaszewski, J. J., & Boyd, T. F. (1953). Colorimetric specification of Munsell repaints. Journal of the Optical Society of America 43, 163-171.

Silverstein, L. D., Lepkowski, J. S., Carter, R. C., & Carter, E. C. (1986, January). Modeling of display color parameters and algorithmic color selection. Paper presented at S.P.I.E. Conference, Los Angeles.

Silverstein, L. D., & Merrifield, R. M. (1985). The development and evaluation of color display systems for airborne applications. Phase I: Fundamental visual, perceptual, and display system considerations. Technical Report DOT/FAA/PM-85-19.

Appendix

Selection of Color Sets

The color sets chosen for study are adaptations and modifications of color sets found in the literature (see Table I). Not all of these authors intended the colors to be used on CRT displays. In some cases the attempt was made to simply find surface colors that were easily discriminable, and rarely confused. We made use of this information only as a starting point to produce color sets for testing in this experiment. The color sets we employed are not exact replications of the colors put forth by these authors. Even if we had desired, it would not have been possible to make exact replications for the following reasons. First, often the luminance information for the colors was not provided. Therefore, although the chromaticity was, in principle, possible to duplicate, the exact color was not. Clearly, two colors of the same chromaticity but very different luminances could appear very different. The importance of brightness contrast between CRT colors on color discriminability has already been demonstrated (Neri, Luria, & Kobus, in press).

Second, and not surprisingly, not all authors recommended sets consisting of exactly seven colors. Therefore we had to make decisions about adding and subtracting colors, decisions that could certainly affect the performance of the sets.

Third, many recommended colors fall outside the color space that could be produced by our CRT. This, of course, was often due to the fact that much of the previous research involved filters or surface colors.

Fourth, we wanted to manipulate most sets in order to achieve greater control over the range of ΔE^* values. In this way we were more likely to perceive the effects of ΔE^* on the dependent measures employed. Therefore, some sets were revised to move the colors apart in chromaticity space in one version, and perhaps move them closer together in another. We were thus more able to examine the importance of factors such as color difference values, the presence or absence of particular colors, and luminance contrast between colors. The particular results we obtained, therefore, can only be interpreted as reflecting the adequacy or inadequacy of our modified versions of the original color sets.

Information about Color Sets Tested

The following is a brief description of the sources of the color sets and the nature of the modifications we made.

Set A. These colors were drawn from a set of 10 used in Luria, et al. (in press). Colors 1-6 and 9 were selected to give a wide range of colors and result in large minimum and mean ΔE^* values.

Set B. These colors were derived from a color set suggested by Laycock (1982) for use in airborne displays. Luminance was not specified. To as great an extent as possible, they fall within six of his seven suggested color regions, plus white. Yellow was omitted. The dominant wavelengths of the colors were modified somewhat in this set to ensure that the minimal ΔE^* value was greater than 40.

Set C. This set was selected from a set of 20 used in Jacobsen (1985). Seven out of the 11 colors showing the fastest mean reaction times were chosen, with an attempt to avoid similar shades and therefore maximize the minimum and mean ΔE^* values.

Set D. These colors were adapted from a set recommended by Meister & Sullivan (1969) and reprinted in Cook (1974). The colors recommended were for surface displays, not CRT displays. Consequently many could not be realized on our monitor. Using the 1931 CIE chromaticity diagram, we brought these colors in along an approximate line from illuminant D65 to the suggested color, until they plotted just inside the triangle representing colors that could be displayed on our CRT. In this way the colors maintained approximately the same dominant wavelength while having the maximum excitation purity that we could produce.

Set E. As in set B, this set was derived from Laycock (1982) but was constructed using generally lower luminances so as to yield lower mean and minimum ΔE^* values. The major chromatic difference was a shift in cyan toward the blue. Yellow was again omitted.

Set F. These colors were adapted from the same source as set D, but the luminances were increased while keeping chromaticity the same. This resulted in an increased mean ΔE^* value, yielding the largest value of any set we tested. The minimum ΔE^* value was not significantly affected.

Set G. These colors are a rough approximation of the subset of eight colors tested by Feallock, Southard, Kobayashi, & Howell (1966) which resulted in no confusions on a color identification task under all the lighting conditions by all the subjects, including deutans. In order to reduce the number to seven we removed the light green.

Set H. These colors are roughly based on a series of ten hues tested by Halsey & Chapanis (1951). Those with dominant wavelengths of 494, 504, and 515 nm were dropped because their approximations on the CRT appeared too similar to the blue and yellow-green that were used. This left seven colors of which the first four had dominant wavelengths equal to those in the original set. We shifted the dominant wavelengths of the last three toward red and purple in order to reduce mean and minimum values of ΔE^*. Luminance values for the seven colors were chosen so that they were approximately equal in brightness, as in the original experiment.

Set I. This set is loosely based on Butler & McKemie's (1974) color code V, consisting of eight colors. It was one of the best two of the seven codes they tested. We dropped the eighth color and approximated the other colors in little more than color name in order to reduce the mean and minimum ΔE^* values. However, the luminance ordering between the colors was maintained, with yellow-green being the dimmest and cyan the brightest.

Set J. This set is loosely based on Butler & McKemie's (1974) color code III, consisting of seven colors. It was the other one of the best two of the seven codes they tested. Again the colors are similar in little more than name. The luminance ordering between the colors was again maintained, with blue being the dimmest and yellow-white the brightest. The chromaticities and luminances were altered to minimize mean and minimum ΔE^* values, while maintaining a set of seven distinguishable colors.

Color Reproduction and Measurement

To produce colors on our CRT as accurately as possible we employed the following procedure. We first measured the luminances of the three phosphors throughout their 8 bit intensity range with a Spectra-Pritchard Model 1980 photometer. A computer program enabled us to use this data to approximate colors by inputting the desired chromaticity coordinates and luminances. In the cases where only the Munsell notation was provided in the literature, the conversion to chromaticity coordinates was made via published tables (Granville & Nickerson, 1943; Kelly, Gibson & Nickerson, 1943; Nickerson, Tomaszewski & Boyd 1953). We chose luminance values that resulted in the desired discriminability between colors, while keeping within the limits prescribed by the color name.

To obtain an accurate documentation of the actual colors employed, all colors were subsequently measured with a Photo Research Fast Spectral Scanner, Model PR-703A that provided both the chromaticity coordinates and luminance values shown in Table IV.

TABLE IV

THE LUMINANCES (C/m^2) AND CHROMATICITY COORDINATES (C.I.E. 1931) OF THE 70 COLORS TESTED IN THIS STUDY.

SET	COLOR	C/m^2	x	y
A	Dark Blue	17.0	.15	.07
	Yellow	189.4	.42	.46
	Red	56.3	.61	.34
	Aqua	85.6	.25	.36
	Purple	19.7	.27	.14
	White	239.7	.29	.30
	Pink	106.3	.35	.33
B	Cyan	81.0	.21	.26
	Yellow-Green	104.7	.30	.54
	Orange	101.8	.50	.41
	Red	63.3	.54	.32
	Purple	68.4	.27	.15
	Blue	28.9	.15	.07
	White	229.8	.28	.31
C	White	183.6	.29	.31
	Dark Green	7.4	.31	.57
	Orange	81.8	.54	.39
	Red	15.7	.49	.27
	Yellow	190.5	.42	.47
	Medium Blue	9.3	.17	.12
	Tan	41.8	.38	.36
D	Green	42.4	.24	.35
	Yellow	75.0	.46	.43
	Amber	62.7	.53	.38
	Red	62.5	.53	.31
	Magenta	85.1	.25	.18
	Blue	59.2	.17	.12
	Gray	68.9	.28	.28
E	Cyan	73.2	.19	.18
	Yellow-Green	62.1	.30	.55
	Orange	59.9	.51	.40
	Red	55.1	.54	.33
	Purple	68.6	.27	.15
	Blue	27.1	.17	.09
	Gray	68.8	.28	.27

TABLE IV continued

SET	COLOR	C/m^2	x	y
F	Green	235.7	.25	.38
	Yellow	140.8	.46	.44
	Amber	92.0	.53	.38
	Red	63.9	.52	.31
	Magenta	132.9	.25	.16
	Blue	80.9	.17	.13
	White	231.4	.29	.31
G	Red	16.7	.61	.32
	Medium Purple	10.3	.25	.17
	Orange-Yellow	82.8	.45	.44
	Dark Yellow-Green	14.5	.31	.53
	Pale Purple-Blue	31.8	.23	.21
	Gray-Red	31.0	.39	.31
	Pale Orange-Yellow	78.0	.33	.34
H	Purple	5.1	.18	.09
	Blue	6.3	.19	.16
	Green-Yellow	4.9	.33	.49
	Yellow	5.6	.40	.39
	Red-Orange	4.2	.50	.35
	Red	4.8	.49	.29
	Mauve	5.2	.41	.25
I	Yellow-Green	0.8	.35	.55
	Magenta	1.1	.24	.13
	Purple-Blue	2.2	.21	.12
	Yellow-Green-White	3.0	.22	.23
	White	10.7	.27	.25
	Yellow	15.9	.43	.40
	Cyan	27.7	.20	.18
J	Blue	0.4	.20	.12
	Yellow-Green	1.0	.40	.49
	Orange	1.2	.46	.43
	Cyan 1	3.1	.22	.20
	Cyan 2	13.0	.22	.23
	Cyan 3	27.5	.21	.18
	Yellow-White	46.3	.35	.35

TABLE V

RESULTS OF NEWMAN-KEULS TESTS FOR THE SEVEN COLOR SETS SHOWING THE SIGNIFICANT DIFFERENCES IN SECONDS BETWEEN COLOR PAIRS.

A

	Dk Blue	White	Aqua	Yellow	Pink	Purple
Red	.047	.086	.149	.183	.194	.204
Dark Blue		.039	.102	.136	.147	.157
White			.063	.097	.108	.118
Aqua				.034	.045	.055
Yellow					.011	.021
Pink						.010

B

	Yel-Gre	White	Orange	Red	Purple	Cyan
Blue	.040	.047	.054	.107	.136*	.156*
Yel-Gr		.007	.014	.067	.096	.116
White			.007	.060	.089	.109
Orange				.053	.082	.102
Red					.029	.049
Purple						.020

D

	Red	Yellow	Magenta	Green	Amber	Gray
Blue	.029	.073	.138	.162	.171	.217*
Red		.044	.109	.133	.142	.188*
Yellow			.065	.089	.098	.144
Magenta				.024	.033	.079
Green					.009	.055
Amber						.046

E

	Yel-Gre	Orange	Purple	Red	Blue	Cyan
Gray	.002	.041	.090	.112	.121	.222*
Yel-Gre		.039	.088	.110	.119	.220*
Orange			.049	.071	.080	.181
Purple				.022	.031	.132
Red					.009	.110
Blue						.101

TABLE V continued

H	Yellow	Blue	Red-Ora	Mauve	Purple	Red
Gre-Yel	.139	.217	.243	.255	.310*	.384**
Yellow		.078	.104	.116	.171	.245
Blue			.026	.038	.093	.167
Red-Ora				.012	.067	.141
Mauve					.055	.129
Purple						.074

I	Yellow	Cyan	Mag	Pur-Blu	Yel-Gr-Wh	White
Yel-Gre	.028	.142	.204*	.292**	.303**	.320**
Yellow		.114	.176	.264**	.275**	.292**
Cyan			.062	.150	.161	.178
Magenta				.088	.099	.116
Pur-Blu					.011	.028
Yel-Cr-Wh						.017

J	Blue	Yel-Gre	Orange	Cyan3	Cyan1	Cyan2
Yel-Whi	.091	.123	.264*	.273*	.432**	.494**
Blue		.032	.173	.182	.341*	.403**
Yel-Gre			.141	.150	.309*	.371**
Orange				.009	.168	.230
Cyan3					.159	.221
Cyan1						.062

* p<.05
** p<.01

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NSMRL Report No. 1068	2. GOVT ACCESSION NO. AD-A163 894	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AN EVALUATION OF COLOR SETS FOR CRT DISPLAYS		5. TYPE OF REPORT & PERIOD COVERED Interim report
		6. PERFORMING ORG. REPORT NUMBER NSMRL Report No.1068
7. AUTHOR(s) NERI, D. F., LT, MSC, USNR JACOBSEN, A. R. LURIA, S. M.		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Submarine Medical Research Laboratory Naval Submarine Base New London Groton, Connecticut 06349-5900		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 65856N M0100.001-1022
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Submarine Medical Research Laboratory Naval Submarine Base New London Groton, Connecticut 06349-5900		12. REPORT DATE 31 Dec 1985
		13. NUMBER OF PAGES 17 plus 7 appendix pages
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Naval Medical Research and Development Command Naval Medical Command, National Capital Region Bethesda, Maryland 20814		15. SECURITY CLASS. (of this report) 'NCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Color-coding; CRT displays; Color-matching; color-difference equations

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Ten subjects performed a color matching task with ten sets of seven colors, adapted from sets recommended in the literature. The color sets covered a wide range in color difference values (ΔE^* in CIELUV, 1976). Performance with some color sets was significantly better than that with others on a task where color discrimination was important and the colors need to be discriminated quickly and accurately. Effectiveness of a color set did not depend on the inclusion of any particular colors; rather, the relationships between colors proved to be important. The larger the color difference value between the two most

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE UNCLASSIFIED

S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. continued:

perceptually similar members of a set, the shorter the reaction time and the fewer the errors on the color matching task.